DOCUMENT RESUME

ED 213 930 CE 031 651

AUTHOR Hines, Constance V.; And Others
TITLE A Study of the Construct Validity of Six Vocational Achievement Tests in the Ohio Vocational Education Achievement Test Program.
PUB DATE [81]
NOTE 20p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; Agricultural Engineering; Auto Mechanics; Carpentry; Distributive Education; Food Service; Home Economics; Secondary Education; State Programs; Test Construction; *Testing Programs; Test Reliability; *Test Validity; *Vocational Education
IDENTIFIERS Fabric Services; Kuder Richardson Reliability Formulas; *Ohio Vocational Education Achievement Test Program

ABSTRACT

A study examined the construct validity of six vocational achievement tests used in the Ohio Vocational Achievement Testing Program. (Subject areas covered in the tests were agricultural mechanics, carpentry, diesel mechanics, distributive education for food services personnel, fabric services, and home economics/food service.) In order to validate the six tests, they were administered to 4,627 tenth, eleventh, and twelfth grade students and job incumbents from a wide cross-section of urban, suburban, and rural locations in Ohio. After computing the Kuder Richardson Formula 20 reliability coefficient for each test to measure its internal consistency, researchers assessed the construct validity of each test through a comparative analysis of the performance levels of the above-mentioned groups on the six tests. A two-factor analysis of variance design, incorporating academic aptitude as a blocking factor, was then employed to investigate group differences in performance levels on each test. Evidence gathered from the study suggests the reliability of each of the six tests and their subtests and further suggests that they could be valuable tools to be used by administrators and teachers as part of a curriculum analysis and inspection system. (MN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED213930

A STUDY OF THE CONSTRUCT VALIDITY OF SIX VOCATIONAL ACHIEVEMENT TESTS IN THE OHIO VOCATIONAL EDUCATION ACHIEVEMENT TEST PROGRAM

CONSTANCE V. HINES, JAMES W. ALTSCHULD AND RAY E. RINDERER

THE OHIO STATE UNIVERSITY

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Constance V. Hines

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Since the late 1950's the Division of Vocational Education of the Ohio Department of Education has been testing the achievement of vocational students in a number of vocational-technical programs in the State of Ohio (The Ohio Vocational Education Achievement Test Program, 1981). This has been accomplished by the staff of the Instructional Materials Laboratory of The Ohio State University. Students are tested in the spring of their junior and senior years.

The use of the Vocational Achievement Tests has expanded beyond Ohio. In 1980 they were administered to approximately 59,000 students in 13 states. The tests are continually being refined and updated in order to reflect the changes in technology in the occupational areas.

Constance V. Hines is a Doctoral Candidate in Educational Foundations and Research, James W. Altschuld is an Assistant Professor in Educational Foundations and Research, and Ray E. Rinderer is the Achievement Test Coordinator of the Instructional Materials Laboratory, The Ohio State University, Columbus, Ohio.

CE031651

The primary purpose of the tests is to provide information for the improvement of instruction.

Given the magnitude and importance of the Ohio Vocational Achievement Testing program, there is a need to systematically analyze the quality of the achievement tests in it. The purpose of this research was to conduct a validity study of six (6) new or completely revised tests in the vocational achievement testing program (Hines & Altschuld, 1980). In order to validate the six tests they were administered in the spring of 1980 to a sample of 10th, 11th, and 12th grade students as well as job incumbents, in the State of Ohio. The comparison of the results of these known groups is a way of demonstrating the construct validity of the tests.

Method

Instrument Development

Each test in a specified vocational area was developed and reviewed by a test development committee consisting of an individual from the state supervisory staff, selected teachers in the occupational area, a representative of the given occupation and a representative of The Ohio State University, Instructional Materials Laboratory.

Each test is designed to measure critical skills and knowledge required for effective job performance as determined from a task analysis of a given occupational area. On the average a test consists of 300-400 multiple choice items (4 alternatives per item) and is organized into a number of subtests (N=7-18) with each subtest assessing student knowledge and understanding of a specific aspect of an occupation.

The six tests included in this validation study--Agricultural Mechanics, Carpentry, Diesel Mechanics, Distributive Education-Food Services Personnel, Fabric Services and Home Economics-Food Services--reflect a diversity of vocational-technical instructional areas. These instruments comprise 20% of the entire test program and were developed by procedures identical to those of the other vocational education achievement tests in the program.

Sample

Teachers who had assisted in the development of the tests as well as other vocational education teachers and administrators were requested to help in the selection of the sample. They nominated schools and intact classes within schools for the study. In addition they identified former students currently employed in the specified occupational field or relevant local employers who in turn supplied the names of job incumbents. Letters were sent to job incumbents explaining the nature of the study and soliciting their voluntary participation in it.

Thus, the non random sample utilized for validation of the six vocational achievement tests consisted of 4,627 subjects from a wide cross-section of urban, suburban and rural locations in Ohio. It included high school sophomores (N=162), vocational program juniors (N=2,156), vocational program seniors (N=2,101) and job incumbents within their first two years on the job (N=208). The sample of students was drawn from comprehensive high schools having vocational programs, vocational schools and secondary feeder schools. Subjects had not been previously exposed to the tests used in this validation study.

Administration of Instruments

A test battery, which included an academic aptitude test (The California Short Form Test of Academic Aptitude, Level 5, CTB/McGraw-Hill, 1970) and The Ohio Vocational Education Achievement Test in a specified occupational area, was administered to a sample of subjects over a consecutive three (3) day period. Total testing time was five (5) hours. The academic aptitude test was administered on the first day, while the vocational achievement test was administered in two parts--Part 1 and Part 2--on the second and third days respectively.

Test monitors were given oral and written instructions for administering the tests, so that a high degree of standardization in administration was achieved.

Procedure

To determine that a necessary prerequisite for test validity was satisfied, the test data for all six vocational achievement tests were subjected to test reliability analyses. For each of the achievement tests, the Kuder Richardson Formula 20 reliability coefficient (KR-20)--a measure of a test's internal consistency--was computed for the total test and each of its subtests.

To assess construct validity, a comparative analysis of the performance levels of four "known groups" on each of the six vocational achievement tests was undertaken (Kerlinger, 1973). The known groups consisted of a sample of high school sophomores (uninstructed in the occupational area in which they were tested), vocational program juniors (one year of instruction in the occupational area), vocational program seniors (two years of instruction in the occupational area), and job incumbents (who

had completed the requisite instructional program or training and were in their first two years on the job). These four known groups therefore differed in the amount of instruction and on-the-job experience which they had received in the occupational area in which they were tested.

It was hypothesized that achievement test scores of the groups would significantly increase as the amount of participation in vocational instruction and training and on-the-job experience increased, if the specific achievement test were a valid measure of the skills and knowledge it purported to measure.

A two-factor analysis of variance design, incorporating academic aptitude as a blocking factor was employed to investigate group (rank) differences in performance levels on each of the six vocational achievement tests. The blocking factor was incorporated into the design in order to take into account the effect of academic aptitude on achievement test scores. By use of the median score for the total sample in a given testing area, aptitude was categorized into two levels--high and low. Follow-up analyses to the analysis of variance (ANOVA) employing Scheffe's post hoc multiple comparison tests and eta-squared coefficients (Kennedy, 1977) were also undertaken.

Results

Test reliability

The test reliability analyses revealed a high degree of internal consistency for the total test as well as subtests for each of the six vocational achievement tests. A summary of total test reliabilities by area is presented in Table 1. An examination of this table indicates

that for each of the six tests, the internal consistency reliability estimate (KR-20) was as high as .97 or greater.

INSERT TABLE 1 ABOUT HERE

The results for the subtest reliabilities given in Table 2 show that in the majority of instances, subtest reliability estimates exceeded .70 (KR-20). Only in 2 of 75 instances was the reliability estimate less than .50. Generally, the lower obtained reliability coefficients tended to cluster in the shorter subtests--not an unexpected result.

INSERT TABLE 2 ABOUT HERE

The observed results described in the two tables collectively indicate that a necessary condition for test validity, i.e. test reliability, was satisfactorily met for each of the six vocational achievement tests.

Test Validity

Academic aptitude and achievement test group means and standard deviations by testing area are presented in Table 3. Academic aptitude test results as would be anticipated increase as a function of rank. An examination of the vocational achievement test results reveal that except in one instance (that of the job incumbents in the Diesel Mechanics test), the pattern of mean performance of the known groups across these six tests was consistent with the proposed a priori hypothesis cited earlier. The analysis of variance results summarized in Table 4 provide further confirmation of these observations.

INSERT TABLES 3 AND 4 ABOUT HERE

No significant effects for the interaction terms (Aptitude X Rank) are observed for any of the six achievement tests. For each of these tests however, the two main effects, academic aptitude and rank (known groups) are significant; with levels of academic aptitude functioning similarly across levels of rank. Subjects with high academic aptitude scores in general, obtained significantly higher achievement test scores than their peers with lower academic aptitude scores. In several tests (Carpentry, Fabric Sciences and Home Economics-Food Services) academic aptitude is explaining a greater percentage of variance than is the variable of concern--known group membership.

The overall F test for rank as noted above was significant for each of the six achievement tests. Follow-up analyses employing Scheffe's post hoc multiple comparison procedure revealed that wherever significant differences in pairwise comparisons were observed, subjects at the higher rank level performed better than those at the lower rank level. Only in a few instances did the observed difference in pairwise comparisons fail to reach the required level of statisitcal significance ($p < .05$). However in all but one of these instances the group at the higher rank level obtained a higher mean score than that obtained for the lower rank level (see Tables 10 and 12 of Hines & Altschuld, 1980, for complete data). The exception is in the Diesel Mechanics test where the seniors outperformed the job incumbents as is clearly indicated in Table 3.

In general, the pattern of results obtained across the six vocational achievement tests is in accord with the group performance patterns

hypothesized, i.e. groups with a greater amount of participation in vocational instruction and on-the-job experience would outperform those with a lesser amount of participation.

A visual inspection of rank subtest scores across all six achievement test, indicates that in general, the same pattern of performance levels across rank was observed at the subtest level (Hines & Altschuld, 1980). Thus, the pattern observed for the total tests was generally maintained for the subtest results.

An examination of the obtained eta-squared coefficients across the six vocational achievement tests revealed that on three of them--Agricultural Mechanics, Diesel Mechanics and Distributive Education-Food Services Personnel--the rank variable accounted for a moderate percentage (10%-23%) of the variance in achievement test scores. Eta-squared coefficients of greater than .10 correspond to a relationship between group membership and achievement test results of at least .30 or higher. While larger eta-squared coefficients would be desirable, .10 could be viewed as a meaningful cutoff point for indicating the strength of underlying relationships between the variables of concern. For the three tests just cited, the results show a moderate to fairly strong relationship between rank membership and vocational achievement test performance. In the case of the Carpentry test, the limited number of ranks available for analysis (N=2) may partially explain the smaller eta-squared value (.06). Thus, approximately 6% of the variance in vocational achievement scores can be accounted for by differences in rank membership between only two groups (seniors and juniors).

For the Fabric Services and Home Economics-Food Services tests, the relationship between rank membership and vocational achievement

test performance was not as strong as in the other tests. With the full complement of four ranks available for analysis for each of these tests, the rank variable explained only 4% and 3% respectively, of the total variance in achievement test scores. A potential reason for the low percentage of explained variance may be obtained by a careful perusal of Table 3. Limited differences in scores of sophomores (uninstructed in the field) and juniors (one year of instruction) are noted in the Fabric Services Tests as well as a rather small variation in the scores of juniors and seniors, given that there is a full year of instruction separating the latter two groups. In Home Economics-Food Services, this observation also holds for the juniors and seniors.

Notwithstanding the relatively few exceptions of areas of weakness identified above, the overall achievement test results in terms of the statistical and substantive significance observed, is in accordance with the pattern of differences in group performance levels hypothesized.

Discussion

The results of the test reliability analyses for the six vocational achievement tests showed a high degree of internal consistency at the total test level and, at the subtest level as well. Thus, a necessary prerequisite for test validity--test reliability--was satisfied for each of the tests under study.

The results of the statistical analyses employed to investigate tests validity suggest as tenable the validity of each of the six achievement tests. In all but one instance the hypothesized pattern of group performance was observed for the descriptive data across all six tests. Generally,

the patterns were further confirmed as statistically significant from the results of the analyses of variance and subsequent follow-up procedures. The sole instance in which there was a shift from the hypothesized pattern of group performance levels for the total test data was observed between the job incumbents and seniors on the Diesel Mechanics test. The seniors obtained a higher mean score ($\bar{X}$=186.8) though not significantly higher, than the job incumbents ($\bar{X}$=179.6). This finding was contrary to the conditions hypothesized for the tenability of test validity. A plausible explanation for this occurrence could be offered in that there were undoubtedly problems encountered in obtaining a sample of job incumbents sufficiently large for validation purposes and representative of the given job incumbent population. In the case of the Diesel Mechanics Achievement test with the incidence of intense specialization in this field, the problem became even more acute. A very small sample of job incumbents would most likely not be sufficiently representative of the many different areas of specialization in the field. Consequently it is probable, that the job incumbent sample obtained was under-represented in certain subtest areas or areas of specialization in the field. It is also likely that job incumbents who were working in one specific area of specialization would have forgotten or had limited knowledge of critical areas of content relevant to another area. The seniors, who were a much larger and more representative sample, had just completed the required instructional program and may have had the advantage of recall of the different content areas assessed by this test.

Notwithstanding these results, support for the validity of this achievement test still seems tenable. With respect to performance comparisons among the other groups on this test, the hypothesized

pattern holds. In addition, an examination of the strength of the relationship between the rank variable and achievement test performance revealed that the former variable explained as much as 13% of the variance in vocational achievement test scores between groups, indicating a moderate relationship between the two variables of interest.

The strength of the relationship between the rank (known groups) factor and vocational achievement test performance as given by the eta-squared coefficient varied somewhat, however, across tests. For four of the tests--Agricultural Mechanics, Carpentry, Diesel Mechanics (as discussed earlier) and Distributive Education-Food Services Personnel--there was a moderate to fairly strong relationship between these two factors. As was observed for the Fabric Services and Home Economics-Food Services Achievement tests, however, this relationship though statistically significant, was relatively weak in terms of substantive or practical significance. The non random sampling of subjects, especially the sophomore group for the Fabric Services Achievement test, could possibly be a major explanatory factor in the reduced strength of the relationship observed for this test. An examination of the academic aptitude data reported in Table 3 shows that the mean aptitude score for the sophomores in the Fabric Services testing area was much higher than was typical of the sophomore group in general and even exceeded that of the juniors and seniors in this area. This was not in keeping with the observed pattern of academic aptitude performance in the other testing areas and with the general expectations for a representative sample of sophomores. There was a reasonably strong relationship between academic aptitude and vocational achievement test performance observed in this study. The exceptionally high ability-level sophomore group could account for

some reduction in the observed variability in achievement test scores between groups. Inadequate sampling could therefore be a major factor in accounting for the reduced strength of the relationship between group membership and achievement test performance observed for this test.

A close examination of the achievement test data presented in Table 3 and the post hoc analyses reveals that with one exception (Diesel Mechanics) job incumbents significantly outperformed the remaining three groups. This observation strongly supports the notion of the validity of a set of achievement tests, designed to measure job related skills and knowledge requisite for successful performance in the respective occupational areas to which the tests relate. The difference in performance levels between the sophomores and the next rank, juniors, was also highly significant, except in the case of the aforementioned Fabric Services test. Again the validity of the tests is being upheld.

Maturation is another possible explanation for higher ranks outperforming lower ones, but it is not a likely causal factor due to the specificity and technical nature of the test items. To account for this factor a subsequent study could be conducted in which non vocationally instructed juniors and seniors would be tested in addition to the groups currently included in the validation effort.

Although the observed differences in vocational achievement performance of juniors and seniors were supportive of test validity, there were instances in which one could reasonably have expected a larger mean score difference, given the fact that one full year of instruction in the occupational program area separates these two groups. One possible rationale for these results could be that there is not a sizeable difference

in content in the two years of the curriculum. This rationale of somewhat weak instructional programs is further supported by the significance of the academic aptitude test results. Aptitude in some instances is accounting for a larger portion of overall test variance then is known group membership. The test outcomes therrfore may be suggestive of areas in the vocational curricula in the schools that require modification in terms of content, instructional techniques and curriculum materials. Since the tests are reflective of the requirements in the given occupational area for which the student is being prepared to work, it seems reasonable that they could be used as part of a curriculum analysis and inspection system. Indeed that is the very purpose for which they have been developed.

In summary, the evidence herein presented supports the reliability of each of the six Ohio Vocational Education Achievement tests and their subtests. Overall the tenability of the test validity proposition was also supported by the analyses already described. With improved sampling of the sophomore and job incumbent groups the obtained results might have been even more supportive of test validity.

As reliable and valid measures the tests are valuable tools for administrators and teachers who are seeking to study and improve vocational-technical education.

References

The California Short Form Test of Academic Aptitude - Level 5. CTB/McGraw Hill; Monterey, California, 1970.

Hines, C. V., & Altschuld, J. W. A study of the validity of six of the ohio vocational education achievement test. Unpublished manuscript, College of Education, The Ohio State University, The Research and Evaluation Consulting Service, October 1980.

Kennedy, J. J. An introduction to the design and analysis of experiments in education and psychology. Washington, D.C.: University Press of America, 1977.

Kenlinger, F. N. Foundations of behavioral research. New York: Holt, Rinehart and Winston, Inc. 1973.

The Ohio Vocational Achievement Test Program. The Ohio State University, Instructional Materials Laboratory, 1981.

Table 1

Summary Results for Total Test Reliability

by Vocational Achievement Test

TEST	No. of Test Items	Obtained Reliability (KR-20)
Agricultural Mechanics	364	.97
Carpentry	346	.98
Diesel Mechanics	379	.97
Distributive Education - Food Service Personnel	345	.99
Fabric Services	391	.97
Home Economics - Food Services	386	.98

Table 2

Subtest Length and Reliabilities

by Vocational Achievement Test

TEST	Range of Obtained Reliability Coefficients (KR-20)	No. of Subtests	Range of Items per Subtest
Agricultural Mechanics	Below .49	1	17
	.50 - .59	1	18
	.60 - .69	6	12 - 15
	.70 - .79	2	23
	.80 and above	4	39 - 66
Carpentry	Below .49	-	-
	.50 - .59	1	16
	.60 - .69	-	-
	.70 - .79	4	15 - 29
	.80 and above	6	24 - 76
Diesel Mechanics	Below .49	1	15
	.50 - .59	1	18
	.60 - .69	2	17 - 25
	.70 - .79	6	17 - 43
	.80 and above	3	33 - 69
Distributive Education - Food Service Personnel	Below .49	-	-
	.50 - .59	-	-
	.60 - .69	-	-
	.70 - .79	2	17 - 35
	.80 and above	10	19 - 57
Fabric Services	Below .49	-	-
	.50 - .59	-	-
	.60 - .69	2	18 - 21
	.70 - .79	6	15 - 39
	.80 and above	7	18 - 41
Home Economics - Food Services	Below .49	-	-
	.50 - .59	-	-
	.60 - .69	-	-
	.70 - .79	3	16 - 21
	.80 and above	7	16 - 80

Table 3

Academic Aptitude and Vocational

Achievement Test Results

AREA TESTED	Rank	N	Aptitude $\bar{X}$	Aptitude S.D.	Achievement $\bar{X}$	Achievement S.D.
Agricultural Mechanics	Sophomores	-[1]	-	-	-	-
	Juniors	305	44.0	13.8	145.8	40.7
	Seniors	259	47.2	13.8	178.6	46.3
	Job Incumbents	29	59.6	10.5	230.3	36.3
Carpentry	Sophomores	29[2]	21.2	26.6	119.6	31.0
	Juniors	1130	49.0	14.1	160.9	49.4
	Seniors	1038	51.4	16.2	194.8	56.6
	Job Incumbents	27[2]	59.5	11.3	227.6	37.9
Diesel Mechanics	Sophomores	68	33.2	26.1	115.5	31.1
	Juniors	286	42.8	20.3	173.6	42.2
	Seniors	282	42.6	20.9	186.8[3]	56.4
	Job Incumbents	52	49.1	16.6	179.6[3]	61.4
Distributive Education - Food Service Personnel	Sophomores	-[1]	-	-	-	-
	Juniors	49	45.5	18.0	97.1	52.9
	Seniors	193	49.2	15.5	119.0	63.0
	Job Incumbents	38	53.7	15.5	210.5	43.3
Fabric Services	Sophomores	27	60.6	13.3	138.2	36.1
	Juniors	138	42.5	15.7	143.0	54.0
	Seniors	91	42.1	16.1	159.2	54.7
	Job Incumbents	36	62.8	9.8	217.3	38.8
Home Economics - Food Services	Sophomores	35	33.2	17.5	132.9	46.6
	Juniors	248	42.1	15.4	158.3	57.1
	Seniors	238	42.4	15.7	173.4	68.2
	Job Incumbents	26	53.3	14.6	228.9	33.2

[1]Obtained sample of insufficient size to report the results.

[2]Obtained sample of sufficient size only for descriptive purposes.

[3]Only obtained reversal of known groups results.

Table 4

Analyses of Variance of Achievement Test

Scores by Rank and Aptitude Level

Area Tested	Source	df	Mean Squared	F	ETA^2
Agricultural Mechanics	Rank (Juniors, Seniors, Job Incumbents)	2	66788.41	48.73*	.10
	Aptitude	1	42066.21	30.69*	
	Rank X Aptitude	2	92.69	0.07	
	Residual Error	587	1370.62		
	Total	592			
Carpentry	Rank (Juniors, Seniors)	1	424915.19	189.02*	.06
	Aptitude	1	1208197.48	537.47*	
	Rank X Aptitude	1	100.91	0.04	
	Residual Error	2164	2247.95		
	Total	2167			
Diesel Mechanics	Rank (Sophomores, Juniors, Seniors, Job Incumbents)	3	85866.63	39.20*	.13
	Aptitude	1	99458.37	45.40*	
	Rank X Aptitude	3	4915.99	2.24	
	Residual Error	680	2190.99		
	Total	687			

CONTINUED

Table 4 Continued

Area Tested	Source	df	Mean Squared	F	ETA^2
Distributive Education - Food Service Personnel	Rank (Juniors, Seniors, Job Incumbents)	2	149455.09	44.10*	.23
	Aptitude	1	21692.29	6.40*	
	Rank X Aptitude	2	3043.99	0.90	
	Residual Error	274	3388.76		
	Total	279			
Fabric Services	Rank (Sophomores, Juniors, Seniors, Job Incumbents)	3	13275.18	7.50**	.04
	Aptitude	1	102049.47	57.67**	
	Rank X Aptitude	3	1249.55	0.71	
	Residual Error	284	1769.39		
	Total	291			
Home Economics - Food Services	Rank (Sophomores, Juniors, Seniors, Job Incumbents)	3	21297.51	6.80**	.03
	Aptitude	1	101230.44	32.20*	
	Rank X Aptitude	3	771.52	0.25	
	Residual Error	539	3086.80		
	Total	546			

* $p < .05$

** $p < .001$